AIRA
AMERICAN IMMUNIZATION
REGISTRY ASSOCIATION

CDC – Endorsed Data Elements Flat File Specification

February 24, 2020

February 24, 2020

CDC-Endorsed Data Element Flat File Extract Specification

Introduction

This specification describes a common file format for CDC-Endorsed Data Elements which could be extracted from an IIS. The scope/quantity of the data extracted by the process is beyond the scope of this document and shall be provided in a separate companion guidance document on a per project basis. The goal of this document is to define a reusable extract format.

Overview

Key steps (refer to Figure 1):

1. Develop/implement process to extract specified data elements
2. Execute extraction process (ex. DBMS queries)
3. Store the extracted Patient data elements and Vaccine data elements into a separate ASCII text files (Tab separated values) with a unique IIS Patient ID used to associate data between them.

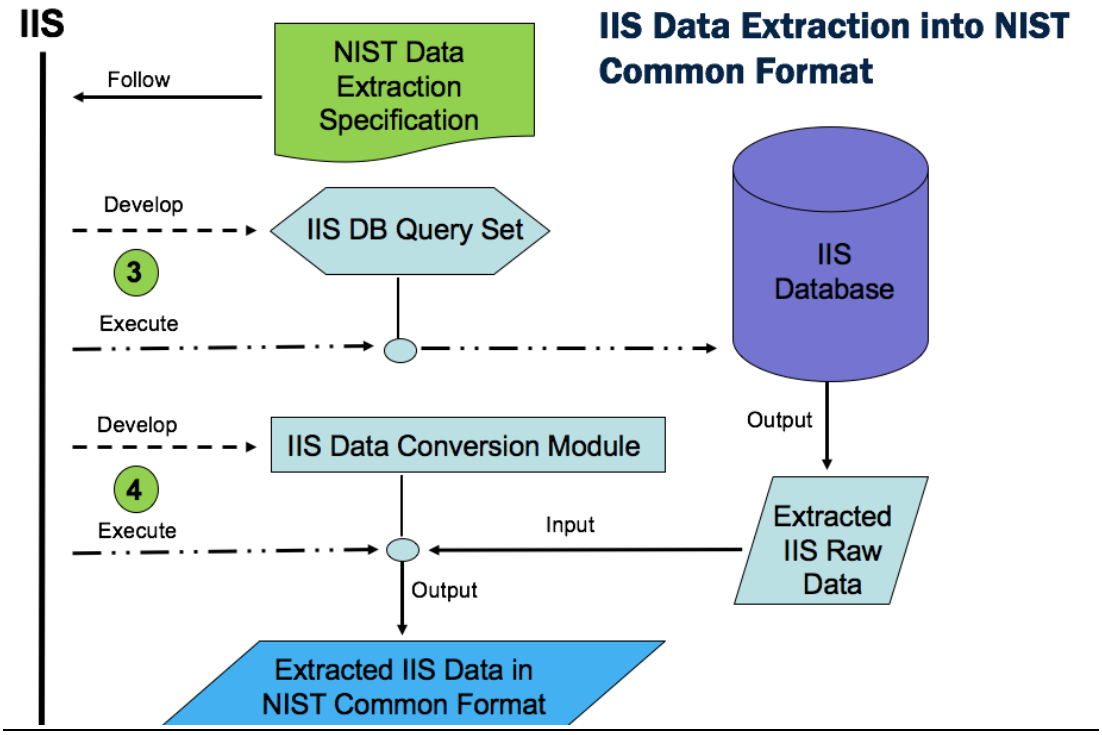

Figure 1: Extract process

Output Format for Extracted Data Elements

All data shall be output to the relevant Extract File as UTF-8 encoded text strings (see: http://wiki.hl7.org/index.php?title=Character_Set_used_in_v2_messages) without change from the local representation except for:

- Elements which represent Dates (e.g. Date of Birth) shall be stored using the date string format: *YYYY-MM-DD (4-character year; 2-character month; 2-character day).*

- Elements represented using a coding system (e.g. Product Type may use CVX codes, local codes, etc.) shall be stored according to the coding system prescribed in HL7 V2.5.1.

Patient and Vaccine Extract File Format

Extracted patient and vaccine-related data elements (see tables 1 & 2) shall be stored in two separate files, a Patient Extract File and a Vaccine Extract File. A Patient ID data element is used to link every patient record to its corresponding vaccine record(s). The structure of both files, referred to below simply as "Extract File," are identical.

Each Extract File is a text file containing lines of tab[1] separated values (TSV), where the first line consists of Data Element Name values from the tables below (Patient or Vaccine) - i.e. the column headings. Each subsequent line represents a logical "record," consisting of tab separated values containing the extracted values corresponding to each "column."

Note that it is possible for some extracted data elements themselves to contain one or more tab (ASCII 09) characters. In such instances, any tab characters *appearing in data elements* must be mapped to space (ASCII 32) characters in the Extract File.

Prior to data extraction, the IIS will be provided a companion document to define which patient and vaccine-related data elements are "in scope" and should be extracted versus those elements which should not be extracted (i.e. excluded). Additionally, prior to data extraction, the IIS shall determine which of the above "in scope" elements may or may not be extracted for any of the following reasons:

- the data element is never collected/stored by the IIS
- local policies prohibit release of the element, and any metadata associated with it (ex. presence, length, etc.)
- local policies only permit release of metadata only

The determination of which data elements shall or shall not be extracted must be consistent across all lines (i.e. records) of the Extract File. Thus, each data element shall either *always* be extracted (and therefore output to the Extract File), or *never* extracted, in which case a predefined string indicator (see Figure 2 below) shall be output to the Extract File in the respective column for that element.

Figure 2 provides a flow chart of how to determine what should be output to the Extract File for any given data element.

[1] "tab" refers to the ASCII HT (Horizontal Tab) character, which has a decimal value of 09, and is the same character typically generated by pressing the 'Tab' key on most keyboards.

Step 1 is for the IIS to consult the companion document to see which elements are "in scope" for the extraction. Elements not in scope shall be represented by the special string "[[EX]]" for excluded (quotes not included) in the Extract File.

Step 2 is for the IIS to determine which of the "in scope" data elements are actually extractable (i.e. are they in the IIS database?). Elements which cannot be extracted because they are not collected and/or not stored in the IIS database shall be represented by the special string "[[NC]]" for Not Collected (quotes not included) in the Extract File.

Step 3 is for the IIS to determine, based on their local policy, which of the remaining data elements (after steps 1 & 2 above) they are *willing* to extract. For example, the IIS may choose *not* to extract the element out of privacy concerns. In that case the element shall be represented by the special string "[[NE]]" for Not Extracted (quotes not included) in the Extract File.

Step 3 (Variation A): Some IIS, however, may determine that although they are not willing to extract the actual values of particular data elements, they *are* willing to indicate whether or not the element is or is not present in the IIS database (such information is essential to the quality analysis for measurement of data "completeness"). In such cases the special strings "[[VP]]" for Value Present and "[[NP]]" for Value Not Present (quotes not included) may represent the element in the Extract File.

Step 3 (Variation B): Similarly, some IIS may determine that although they are not willing to extract the actual values of a particular data element, they *are* willing to extract its length (e.g. it may be useful to the quality analysis to know the length of some data elements such as Patient Name). In such cases the special string "[[LEN {n}]]" (quotes not included) shall be used to represent the data element in the Extract File, where {n} is an integer representing the string length of element's value in the IIS database. Note that data elements which contain empty/null (ex. note present) values in the database should result in {n} being set to 0.

Note that the above cases where data elements *are not* extracted is different from the cases where data elements *are* being extracted but contain empty/null (ex. not

present) values in the IIS database. The latter shall result in <tab><tab> being output to the Extract File.

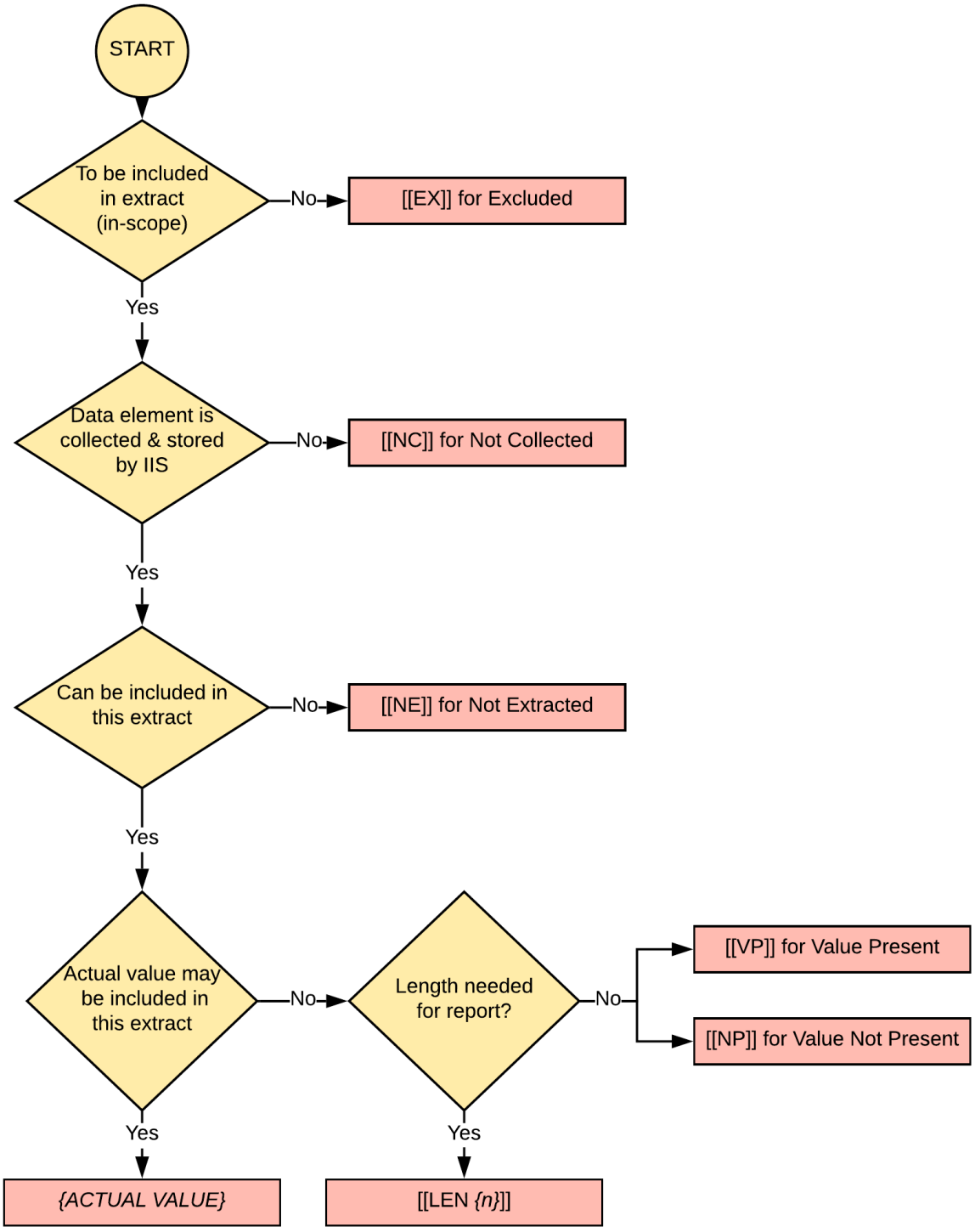

Figure 2: Extraction Decision Tree

Required Elements

Only data elements necessary to link records together (e.g. IIS Patient ID) or are used for report creation (e.g. Reporting Group) are required to always be populated. These are noted in bold. All other fields should be populated or left empty as indicated by the rules above and the directions for this particular export.

The linking elements are required to be unique between all associated files and for the purposes of this extract. However, they are not required to represent the actual identifier, as used by the IIS. They may be generated for one-time use during the extraction. In other words, a record linked with IIS Patient Id “100” may or may not represent the same record in an export at a later date with the same IIS Patient Id “100”. The linking elements are only required to be unique and consistent across a single set of files extracted and submitted at the same time. Thus, extractors have the option of further obfuscating the original patient id by not being required to export usable patient identifiers.

For all other fields, extractors must not skip records that are missing data that would normally be required by an IIS or ensure that empty values are populated with contrived information. This specification is designed to allow extractors to represent the data as it is, with the least amount of translation.

Patient Data Elements

The patient-related data elements to extract to the Patient Extract File are listed in the table below.

Note that all of the patient records in the Patient Extract File shall have at least one record in the Vaccine Extract File (see next section) unless the patient has no immunizations. Thus, if a patient record does not have a corresponding immunization record the analysis will assume this patient has received no immunizations.

Also note that patient information is assumed to be “consolidated” by the IIS prior to extraction, and thus a single set of the elements listed below are expected to be extracted for each Patient. Some IIS may store multiple representations of the same patient record or vaccination event, care should be taken to extract the single best representation (as the IIS can determine it) of the elements below. IIS should not extract all versions and variations of received and stored records.

Table 1: Patient Data Elements

Field #	Data Element Name	Format	Details	Value Set
1	**IIS Patient ID**	String	Unique ID for this patient within this extract file. This may or may not be the actual IIS Patient ID within the IIS DB. A unique value is required within this extract file.	
2	Patient Name - First	String	The patient's first name.	
3	Patient Name – Middle	String	The patient's middle name.	
4	Patient Name - Last	String	The patient's last name.	
5	Mother's Maiden Name	String	The last name under which the mother was born (i.e., before marriage).	
6	Mother's Name - First	String	The first name of the patient's mother.	
7	Mother's Name - Middle	String	The middle name of the patient's mother.	
8	Mother's Name - Last	String	The last name of the patient's mother.	
9	Patient Date of Birth	YYYY-MM-DD	The patient's date of birth.	
10	Patient Gender	String	The patient's gender.	User Defined table 0001 (M, F, U)
11	Patient Address - Street	String	The street component of the patient's address.	
12	Patient Address - City	String	The city component of the patient's address.	
13	Patient Address - State	String	The state component of the patient's address.	US Postal Service two-character Sate Codes

14	Patient Address - Country	String	The country component of the patient's address.	HL70399
15	Patient Address – Zip/postal	String	The zip code of the patient's address.	
16	Patient Race	String	The patient’s race.	User Defined Table 0005
17	Patient Ethnicity	String	The ancestry of the patient.	User Defined Table 0189
18	Patient telephone number	String	The patient’s phone number.	
19	Patient Email address	String	The patient’s email address.	
20	Patient Primary Language	String	The patient’s primary language	User-Defined Table 0296
21	Patient Alias Name: First	String	The first name of a patient's alternate or also-known-as name.	
22	Patient Alias Name: Middle	String	The middle name of a patient's alternate or also-known-as name.	
23	Patient Alias Name: Last	String	The last name of a patient's alternate or also-known-as name.	
24	Responsible Person Name: First	String	The first name of the person responsible for the patient.	
25	Responsible Person Name: Middle,	String	The middle name of the person responsible for the patient.	
26	Responsible Person Name: Last	String	The last name of the person responsible for the patient.	
27	Responsible Person Name: Relationship to Patient	String	The actual personal relationship that the responsible person has to the patient.	User-Defined Table 0063

28	Patient Multiple Birth Indicator	String	Whether the patient was part of a multiple birth (yes/no).	Y, N, or empty
29	Patient Birth Order	String	When a patient was part of a multiple birth a number is defined in this element (e.g., 1, 2, 3).	
30	Patient Birth State	String	State of the birthing facility location.	US Postal Service two-character Sate Codes
31	Patient status indicator—provider level	String	The current active/inactive status of the patient in relation to the provider organization.	User-defined table 0441
32	Patient status—jurisdiction level	String	The patient/geographic jurisdiction status determined from address information when the patient's address within the jurisdiction is known.	User-Defined Table 0441
33	Record Creation Date	YYYY-MM-DD	The date this patient was created in the IIS.	
34	Record Update Date	YYYY-MM-DD	The date this patient was last updated in the IIS.	

Vaccine Data Elements

The vaccine-related data elements to extract to the Vaccine Extract File are listed in the table below.

Note that all immunization records in the Vaccine Extract File shall have a corresponding patient record in the Patient Extract File (see section above). If an immunization record does not have a corresponding patient record in the immunizations will be noted and discarded from analysis.

Table 2: Vaccination Data Elements

Field #	Data Element Name	Format	Details	Value Set
1	**IIS Patient ID**	String	This is the IIS Patient ID from the Patient Extract File.	
2	**IIS Vaccination Event ID**	String	Unique ID for this vaccination event within this extract file. This may or may not be the actual IIS Vaccination Event ID within the IIS Database. A unique value is required within this extract file.	
3	**Reporting Group**	String	This field will be used to group immunization records for reporting purposes. This value is at the discretion of the IIS to assign a meaningful value depending upon how granular the IIS would prefer to see reports. More detail on this is found after this table.	
4	Sending Organization	String	The identifier of the organization that connects to the IIS and submits the record. This may or may not be the same as the Responsible Organization or Administered At Location.	

5	Responsible Organization	String	The identifier of the organization that originated and is accountable for the content of the record. May be referenced as IIS-AO ID. This may or may not be the same as the Sending or Administered at Location.	
6	Administered at Location	String	The facility name/identifier of the facility that administered the immunization.	
7	Vaccine Administering Provider	String	The name of the provider (person) administering the vaccination. Note that this "provider name" could include an actual name, or an ID	
8	Vaccine Type (CVX)	String	The vaccine type of the vaccination event.	CVX
9	Vaccine Type (NDC)	String	The vaccine type of the vaccination event.	NDC
10	Vaccine Administration Date	YYYY-MM-DD	The date the vaccination event occurred.	
11	Vaccine Manufacturer	String	The manufacturer of the vaccine that was administered.	MVX
12	Vaccine Lot Number	String	The lot number of the vaccine administered.	
13	Vaccination Event Record Type	String	Indicates whether the vaccination event is based on a historical record or was given by the administered at location.	NIP001
14	Vaccine Route of Administration	String	The route of administration	HL7-defined table 0162 or NCIT
15	Vaccine Site of Administration	String	The anatomical site where the vaccine was given.	HL7-defined table 0163
16	Vaccine Expiration Date	YYYY-MM-DD	The expiration date of the vaccine administered.	
17	Vaccine Dose Volume	String	The amount of vaccine administered (e.g., .5). Note all	

			volumes will be assumed to be mL. If mL are not captured in the IIS, then this field should be set to [[NC]] for all records.	
18	Ordering Provider	String	The identifier of the provider (person) ordering the immunization. Note: this "provider name" could include an actual name, or an ID. May also be referenced as the prescriber ordering provider Identifier.	
19	Completion Status	String	Indicates if the vaccination event was successfully administered.	HL7-defined table 0322
20	Dose Level Eligibility	String	The program that should pay for a given immunization, based on the characteristics of the patient and the type of vaccine administered.	User defined table 0064
21	Vaccine Funding Source (Dose Level Public/Private Indicator)	String	The funding source of the vaccine administered. That is, was the vaccine administered purchased as publicly funded, privately funded, or with other jurisdiction specific funding.	PHVS_ImmunizationFundingSource_IIS (2.16.840.1.114222.4.11.3287) or locally defined codes
22	VIS - Identifier	String	The VIS GDTI document code	VIS GDTI
23	VIS - Publication Date	YYYY-MM-DD	The publication date of the Vaccine Information Statement.	
24	VIS – Given Date	YYYY-MM-DD	The date the VIS information was provided to the patient.	
25	Record Creation Date	YYYY-MM-DD	The date this vaccination event was created in the IIS.	
26	Record Update Date	YYYY-MM-DD	The date this vaccination event was last updated in the IIS.	

Reporting Group can be used to stratify the data into like groupings. This field and expectations can be further defined in project specific extract specifications. Below are a few examples of how immunization might be grouped:

- Responsible Organization
- Pharmacy submitted vs. non-pharmacy submitted
- Private vs. Public providers
- EHR Vendor system

Each example grouping above could be used to detect different trends in the immunization data.